JIAYAO ZHANG

Philadelphia, PA 19104, US
jiayaozhang[at]{acm,ieee}.org ◇ https://jiayao-zhang.com

RESEARCH SUMMARY

I work on the intersection between machine learning, statistics and natural language processing, on a wide variety of problems, in an effort to answer **how to design novel solutions to NLP problems from rigorous statistical principles and methodologies** (e.g., *how to perform data attribution for text generation models*) and **how to leverage language modeling and related techniques to facilitate data science research** (e.g., *how to perform causal inference better with textual data?*).

EDUCATION

University of Pennsylvania, PA, US — Sept 2019 - May 2024 (Expected)
Ph.D. (Applied Math and Computer Science) M.A. (Dual Masters in Statistics)
Advisors: Dan Roth, Weijie J. Su

University of Hong Kong, Hong Kong SAR, China — Sept 2015 - June 2019
B.Eng. in Computer Science (First Class Honours)

EXPERIENCES

Applied Scientist Intern, AWS AI Labs, Amazon — May 2023 - Aug 2023
Machine Learning & Forecasting Team (led by Dr. Bernie Wang)

- Work on massive datasets, `Python/C++` pipelines
- A holistic study of language modeling on compression [P4]

Research Intern, AllenNLP Team, Allen Institute for AI — Feb 2023 - May 2023
Mentored by Dr. Yanai Elzar and Dr. Jesse Dodge

- Explainability for text generation models via influence functions [P5]
- Causal effect of early arXiv'ing in peer-review [P7]

Applied Scientist Intern, AWS AI Labs, Amazon — May 2022 - Aug 2022 (Full-time), Aug 2022 - Jan 2023 (Part-time)
Machine Learning & Forecasting Team (led by Dr. Bernie Wang)

- Methods for reverse causal inference on panel data [W2]

Visiting Student Researcher, EECS, UC Berkeley — July 2018 - Sept 2018
Working with Prof. Bo Li and Prof. Dawn Song

- Information-theoretic capacity in neural nets [W1]

TEACHING

Teaching Assistant, University of Pennsylvania — 2020 - 2021

- MATH648/STAT930 Graduate Probability (Prof. Robin Pemantle)
- MATH649/STAT931 Graduate Stochastic Processes (Prof. Xin Sun)

PREPRINTS

[P7] Yanai Elazar*, **Jiayao Zhang***, David Wadden*, Bo Zhang, and Noah A. Smith. "Estimating the Causal Effect of Early ArXiving on Paper Acceptance". *Technical Report*. 2023.

[P6] Buxin Su, **Jiayao Zhang**, Natalie Collina, Aaron Roth, Xiao-Li Meng, Dingdong Li, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, and Weijie J. Su. "ICML OpenRank Experiment". [Online] Project Website. 2023.

[P5] **Jiayao Zhang** and TBD. "Data Attribution for Text Generation Models via Approximate Influence Functions". *Work in progess*. 2023.

[P4] **Jiayao Zhang** and TBD. "Lossless Compression using Language Models: Initial Results, Challenges and Opportunities". *Work in progess*. 2023.

[P3] Chen Chang*, **Jiayao Zhang***, Dan Roth, Ting Ye, and Bo Zhang. "Association Between Author Metadata and Acceptance: A Feature-Rich, Matched Observational Study of a Corpus of ICLR Submissions Between 2017-2022". *Technical Report*. 2022.

[P2] **Jiayao Zhang**, Hongming Zhang, Zhun Deng, and Dan Roth. "Investigating Fairness Disparities in Peer Review: A Language Model Enhanced Approach". *Technical Report*. 2022.

[P1] **Jiayao Zhang**, Guangxu Zhu, Robert W. Heath, and Kaibin Huang. "Grassmannian Learning: Embedding Geometry Awareness in Shallow and Deep Learning". *Technical Report*. 2018.

CONFERENCE PROCEEDINGS

[C6] Zhun Deng, **Jiayao Zhang**, Linjun Zhang, Ting Ye, Yates Coley, Weijie J. Su, and James Zou. "FIFA: Making Fairness More Generalizable in Classifiers Trained on Imbalanced Data". *International Conference on Learning Representations (ICLR)*. 2023.

[C5] Zhaowei Wang, Quyet V. Do, Hongming Zhang, **Jiayao Zhang**, Weiqi Wang, Tianqing Fang, Yangqiu Song, and Simon Wong Ginny Y. and See. "COLA: Contextualized Commonsense Causal Reasoning from the Causal Inference Perspective". *Proceedings of the Association for Computational Linguistics (ACL)*. 2023.

[C4] Bo Zhang and **Jiayao Zhang**. "Some Reflections on Drawing Causal Inference using Textual Data: Parallels Between Human Subjects and Organized Texts". *First Conference on Causal Learning and Reasoning (CLeaR)*. 2022.

[C3] **Jiayao Zhang**, Hongming Zhang, Weijie J. Su, and Dan Roth. "ROCK: Causal Inference Principles for Reasoning about Commonsense Causality". *International Conference on Machine Learning (ICML)*. 2022.

[C2] **Jiayao Zhang**, Hua Wang, and Weijie J. Su. "Imitating Deep Learning Dynamics via Locally Elastic Stochastic Differential Equations". *Advances in Neural Information Processing Systems (NeurIPS)*. 2021.

[C1] **Jiayao Zhang** and Anirbit Mukherjee. "Stochastic Oracles for RMSProp with Provable Convergence". *DeepMath Conference*. 2020.

JOURNAL ARTICLES

[J1] Yuqing Du, Guangxu Zhu, **Jiayao Zhang**, and Kaibin Huang. "Automatic Recognition of Space-Time Constellations by Learning on the Grassmann Manifold". *IEEE Transactions on Signal Processing* (2018).

WORKSHOPS

[W2] **Jiayao Zhang**, Youngsuk Park, Hilaf Hassen, Dannielle C. Maddix, Dan Roth, and Bernie Wang. "Towards Reverse Causal Inference: Precise Formulation and Chllenges". *Advances in Neural Information Processing Systems (NeurIPS). Workshop on Causal Dynamical Systems (CDS)*. 2022.

[W1] **Jiayao Zhang**, Ruoxi Jia, Bo Li, and Dawn Song. "On the Weak Neural Dependence Phenomenon". *Advances in Neural Information Processing Systems (NeurIPS). Workshop on Integration of Deep Learning Theories*. 2018.

TALKS

[T7] Jiayao Zhang. "Presentation on FIFA Paper (Virtual)". ICLR 2021, May 1, 2023.

[T6] Jiayao Zhang. "Poster on Commonsense Causality Reasoning". MASC-SLL 2022, Apr. 30, 2022.

[T5] Jiayao Zhang. "Presentation on ROCK Paper (Virtual)". ICML 2022, July 19, 2022.

[T4] Jiayao Zhang. "Talk on Causal Inference Applications in NLP". UPenn Statistics Ph.D. Student Seminar, May 11, 2022.

[T3] Jiayao Zhang. "Talk on Commonsense Causality Reasoning". HKUST KnowComp Research Seminar, Nov. 24, 2022.

[T2] Jiayao Zhang. "Talk on Ph.D. Research". HKUCS Research Seminar, Nov. 10, 2022.

[T1] Jiayao Zhang. "Presentation on LE-SDE Paper (Virtual)". NeurIPS 2021, Dec. 7, 2021.

SOFTWARE

pyTracer: Physically Based Rendering/Ray Tracing `github:zjiayao/pyTracer`
`Python/C++`, Measured BRDF, MC Samping, MH Light Transport

jmt::matrix Linear Algebra Library for Matrix Decomposition `github:zjiayao/jmt-matrix`
`C++`, Real/Complex SVD/QR/Cholesky Decomposition

HONOURS AND AWARDS

Benjamin Franklin Fellowship 2019-Present
Supports Ph.D. research.

Fisherman Ph.D. Fellowship 2019-2020
Supports Ph.D. research.

HKU URFP Best Poster Award April 2019
Awarded for the poster on weak neural dependence.

HKU Chui's Student Excellence Scheme July 2018
Support visit to UC Berkeley (US$ 1,000).

HKU Undergraduate Research Fellowship Program (URFP) 2018-2019
Nurturing next generation of researcher (US$ 5,000).

HKU Foundation/HKSAR Government Scholarship 2015-2019
Annually renewable (US$ 11,852 pa).

Dean's Honours List 2015-2019
Award to top students in the faculty.

Walter Brown Memorial Prizes in Mathematics 2015-2016
Award to the best first-year engineering student in math.

HKSAR Government Reaching Out Awards 2015-2016
Support outbound academic exchange (US$ 1,200).

SERVICES

Student Member
ACM (2017-Present), IEEE (2017-Present), AMS (2019-Present), IMS (2019-Present), FSF (2022-Present)

Reviewer April 2019-Present
IEEE Trans. Signal Processing, *ICML 2022*, *NAACL 2022*, *NeurIPS 2022*, *ICML 2023*, *NeurIPS 2023*

HKU Student Ambassador 2016-2019
"We are the faces of HKU." https://aal.hku.hk/student/ambassador/

Team Leader, Go2Africa (Ghana) Project, Beyond The Pivot 2015-2016
One-month latrine construction and teaching in Ghana. https://hibtp.org

SKILLS

Programming C/C++, Python science stack, LaTeX, SQL, Linux Server, AWS, MatLab, Java

Language Chinese (Mandarin, Cantonese), English, French (Basic), Japanese (Basic)

Last updated in October 2023.